The Bell System Technical Journal
Vol. 53, No. 2, February 1974
Printed in U.S.A.

Peak-Load Traffic Administration of a Rural Multiplexer with Concentration

By S. B. GERSHWIN, R. V. LAUE, and ERIC WOLMAN

(Manuscript received June 19, 1973)

A procedure is proposed for estimating the main-station capacity of an SLM* *(Subscriber Loop Multiplexer) system by observing the traffic load when the system is partially filled. The procedure is intended to be usable in unattended offices, and requires only one measurement per week and very few calculations. In contrast to the usual practice of measuring load in a time-consistent busy-hour, we work with weekly peak loads, and so our method is based upon the statistical theory of extreme values. The validity and precision of the procedure have been investigated by applying it to data from a study of rural traffic and by a Monte Carlo study of its behavior. Use of this administrative procedure should give the average* SLM *system a capacity of about 120 rural residential customers, in contrast to the limit of 80 that would be necessary in the absence of traffic measurements.*

The technique described in this paper was developed for the SLM *system and could be used, with suitable changes of numerical values, to handle any subscriber system with concentration. We also hope that, with some modification, the method will be applicable to the administration of other traffic-carrying systems.*

CONTENTS

* Trademark of the Bell System.

I. INTRODUCTION

The SLM (Subscriber Loop Multiplexer) system is a digital carrier and switching system that was developed to provide economically for main-station growth and upgraded service on long rural cable routes. It is capable of serving 80 lines, all sharing 24 channels. Each of the 80 lines can be used for single- or multi-party service. (For a detailed description of the SLM system see Ref. 1.) For the purposes of this paper, which is concerned with traffic, the SLM system may be viewed as a remote line-concentrator serving 80 lines on 24 full-access channels.

The quality of service given to SLM subscribers should be kept well above levels that might lead to complaints, and to the need for hasty rearrangements that would interfere with the orderly growth of subscriber plant. Hence, service for these subscribers should not be noticeably different from that for customers served by physical pairs to the central office. This service objective will be met if blocking exceeds one-half percent in no more than a few hours per year. (It is possible to imagine a distinct service, for sparsely populated rural areas, in which a less stringent service objective would be appropriate.)

The Rural Line Study,[2,3] a study of subscriber line usage in rural areas, has confirmed that rural residential subscribers like those studied in the territory of South Central Bell can almost always be served on one SLM system with essentially no blocking in groups of 80 main stations or more. (In fact, as shown below, most rural systems should be able to serve many more than 80 main stations.) However, the load per main station does vary greatly from place to place and from customer to customer. Thus even 80 main stations will in a few cases generate enough load to cause undesirably frequent blocking in excess of one-half percent. Some means of monitoring the traffic performance of SLM systems are therefore necessary.

One such means is a register which records the total amount of a system's all-channels-busy time since the register was last read and reset to zero. But such a register, which can indicate by its readings when a system is overloaded, cannot be used to foresee an overloaded condition until too many lines have already been assigned to the system. When the register's readings exceed a specified threshold, the administrator's response must be to remove lines from the system and to serve them on other facilities.* Because of the long lead-times involved in cable planning and installation, a major goal of the work reported here was to avoid this situation.

* Administrative use of this register is described elsewhere.

Since some additional traffic-monitoring capability was necessary, it seemed best to provide a measurement which could be used to *predict* the ultimate main-station capacity of a system, and thus to guide the loading of the system and the planning of relief facilities. The natural quantity to measure is carried load; and the system's second traffic register, which was also chosen to minimize the volume of data to be taken and processed, records peak hourly carried load—that is, the highest hourly carried load that has occurred since this register was last read and reset to zero. As shown below, it is sufficient to take readings once a week. This frequency is compatible with a normal schedule of visits to unattended offices. Note that the time that this peak traffic occurred is neither recorded nor of importance to the procedure to be described, so that the usual time-consistent busy-hour is not identified.

This paper describes a simple system-loading procedure, requiring few measurements and calculations, which should assure that the fraction of calls blocked exceeds 0.5 percent in no more than a few hours during a year. Load measurements are taken on a partially filled system, and from these an estimate is made of the total number of main stations, in the same locale, that the SLM system can safely serve.

This method could be used to administer any subscriber concentrator for which peak-load traffic data can be obtained. Changes in the numbers of lines served or voice channels, in the access afforded by the switching network, or in the handling of intra-system calls would of course require modification of the numerical values used. (And additional problems can arise: For example, partial access may give rise to a need for load-balancing procedures.)

II. THE BASIC PROCEDURE

Weekly peak load measurements can be used whenever 40 or more main stations are assigned to an SLM system. We begin with N weekly peaks, $x_1, \cdots, x_N$. (N is normally equal to 4.) We simply calculate the mean,

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i, \tag{1}$$

and the variance,

$$v = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2, \tag{2}$$

of these peaks. On a chart (see Fig. 1) corresponding to the number of

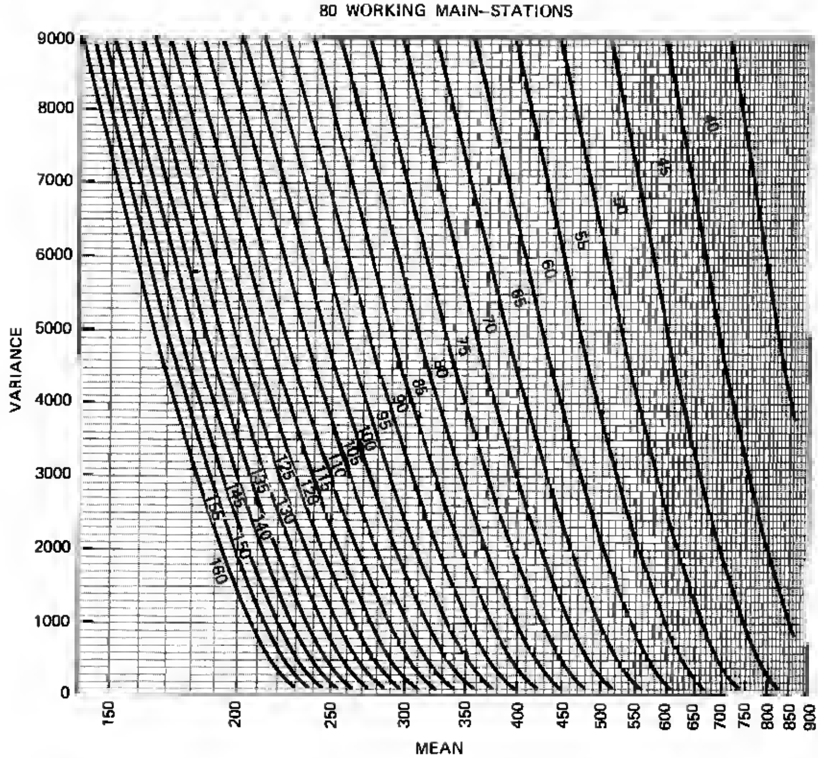

Fig. 1—Estimated main-station capacity as determined by mean and variance of weekly peak loads.

currently working main stations, we find the point defined hy the coordinates $\bar{x}$ and v. This point will fall in one of the regions labeled 40, 45, $\cdots$, 160. Tbe label of that region is the estimated main-station capacity of the SLM system.

For example, let us say we have 80 working main stations. We ohserve four weekly peaks and calculate a mean of 260 CCS and a variance of 1400 $(\mathrm{CCS})^2$. Figure 1 is the chart corresponding to 80 working main stations. The point (260, 1400) falls in the region labeled 120. This is the *estimated* main-station capacity for the system in question. Repeating this estimation procedure four times (using data from 16 weeks of operation) and calculating the weighted mean of the estimates, using the respective numhers of working main stations as weights, we ohtain the *predicted* capacity of the system in

main stations. Some precautions which must be observed in drawing conclusions from this process are summarized below.

III. THE DISTRIBUTION OF WEEKLY PEAK TRAFFIC LOADS

Maynard[3] showed that weekly peak traffic loads of potential SLM subscribers seemed to behave as if they came from an extreme-value distribution of the form

$$G(x) = \exp\left(-e^{-\alpha(x-u)}\right), \tag{3}$$

which is sometimes referred to as "Gumbel's first asymptotic distribution." It has been shown[4] that, for the so-called "exponential" class of distributions, the largest value of a random sample of size n will be asymptotically distributed (as $n \to \infty$) according to (3). The exponential class includes most well-known distributions with an infinite tail to the right, such as the normal, lognormal, and gamma.

The mean of the distribution (3) is

$$E(x) = u + \gamma/\alpha, \tag{4}$$

where $\gamma(=0.5772\cdots)$ is Euler's constant, and the variance is

$$V(x) = \frac{\pi^2}{6\alpha^2}. \tag{5}$$

Gumbel suggests that u and α be estimated by replacing $E(x)$ and $V(x)$ by their sample values, (1) and (2) respectively, and solving (5) and then (4) for α and u:[4]

$$\hat{\alpha} = \frac{\pi}{\sqrt{6v}}, \tag{6}$$

$$\hat{u} = \bar{x} - \gamma/\hat{\alpha}. \tag{7}$$

IV. THE MATHEMATICAL MODEL

With J main stations served by an SLM system, the weekly peaks will have an extreme-value distribution with parameters u_J and α_J. If we increase the number of main stations from J to K, the weekly peaks will have a new extreme-value distribution with parameters u_K and α_K. In this section we describe a method for estimating u_K and α_K when J and K are known and u_J and α_J have been estimated from measurements. That is, we want to know what the distribution of weekly peaks will look like for K main stations when we have observed this distribution with only J main stations being served.

Suppose that, during any week, there are n hours in which the weekly peak traffic load may occur. We know from experience that the weekly peak can occur during almost any waking hour.[3] However, for any given week, n will he much smaller than the numher of waking hours. (In Appendix A we show that $n = 10$ seems to he an appropriate choice for our purposes.) We call these n hours (whose actual times of occurrence are not specified) the *candidate* husy-hours.

We now assume that each main station generates a load with mean μ and standard deviation σ during each of the candidate husy-hours. If customers hehave independently, the load distrihution in candidate husy-hours must have mean $J\mu$ and standard deviation $\sigma\sqrt{J}$. Let this distrihution be F, with density $f = F'$. The weekly peak will then he the maximum value in a random sample of size n from the distrihution F. If F is in the exponential class of distrihutions, the distrihution of this maximum can he approximated hy the extreme-value distrihution (3). Gumhel[4] shows that the parameters u and α are given approximately hy

$$F(u) = 1 - \frac{1}{n} \tag{8}$$

and

$$\alpha = nf(u). \tag{9}$$

Since the candidate-husy-hour loads are the sums of the loads from J main stations, it seems reasonahle to assume that F is normal,* as suggested hy the central-limit theorem. (As mentioned ahove, we take J and K to he at least 40.)

Let Φ he the standard unit-normal distrihution function and $\phi = \Phi'$ the corresponding density. Define ν hy the relation

$$\Phi(\nu) = 1 - \frac{1}{n}. \tag{10}$$

Then ν is the $1 - (1/n)$ quantile of the unit-normal distrihution, for which tahles and computer suhroutines are availahle. Then from (8) and (9) it is readily seen that

$$u_J = J\mu + \nu\sigma\sqrt{J} \tag{11}$$

and

$$\alpha_J = \frac{n\phi(\nu)}{\sigma\sqrt{J}}. \tag{12}$$

* The gamma distribution was also considered. A study which led to the choice of the form of the candidate-busy-hour load distribution, and to the number n of candidate busy-hours in a week, is described in Appendix A.

If the $K - J$ subscribers to be added come from the same population as the J subscribers already being served, the candidate-busy-hour load distribution for K main stations will be normal with mean $K\mu$ and standard deviation $\sigma\sqrt{K}$. The weekly-peak-load distribution will have parameters defined by (11) and (12) with J replaced by K. From the four equations (11), (12), and the corresponding equations for K main stations, the variables μ and σ can be algebraically eliminated to yield these expressions for u_K and α_K in terms of u_J and α_J:

$$u_K = \frac{K}{J} u_J - \left[\frac{K}{J} - \sqrt{\frac{K}{J}} \right] \frac{C_n}{\alpha_J}, \tag{13}$$

$$\alpha_K = \sqrt{\frac{J}{K}} \alpha_J. \tag{14}$$

Here $C_n = n\nu\phi(\nu)$, a function of n only. Hence, for a given n, we can estimate the parameters of the weekly-peak-load distribution for K main stations by observing the weekly peaks generated by $J(<K)$ main stations.

We can choose K so that the weekly-peak-load distribution defined by u_K and α_K is such that the system satisfies the service criterion (which is described below). This entire calculation can be incorporated in a series of charts, each corresponding to a different value of J. An example is that given in Fig. 1 for $J = 80$.

V. THE SERVICE CRITERION

To introduce the service criterion, we define a *heavy-load hour* as an hour in which the offered load is such that the probability of blocking is 0.005 or greater. We first attempted to limit the frequency of such hours to no more than four times a year, but found, as shown below, that the statistical characteristics of our administrative procedure lead to a different formulation of the service criterion.

Jones proposed a model for relating the probability of blocking in a concentrator to the number of channels, the number of customer lines, the percentage of intra-system traffic, and the total source load.[5] Johnson has written a computer program which calculates load-service relations based on Jones's model.[6] Although Jones's model assumes blocked calls cleared, whereas waiting and retrials occur in a real system, we believe that these effects are compensated by the lopsided distribution of load per line,[7] so that this model is appropriate. Laue showed from the Rural Line Study that, for 80 main stations, the expected intra-system traffic (IST) should be about 18 percent,

but because of temporal and customer variation we used the more conservative value 30 percent.[2] (In subscriber concentrators which are not arranged for remote switching of intra-system calls,* such calls occupy two channels. Thus for a given offered load, IST increases the variability of the traffic and hence the blocking also.) It is known that the percentage of IST increases with the number of main stations; but the value 30 percent is used throughout because the smoothing effect of party-line interference, which we neglect, grows with the number of main stations.[2] Omission of party-line interference from the model is equivalent to treating a main station as a "traffic source." For 24 channels and an IST of 30 percent, the finite-source effect makes that load which results in a 0.005 probability of blocking a decreasing function of the number of main stations. This load varies from 526 CCS for 40 main stations to 471 CCS for 160 main stations; we call it $L(K)$, the load that causes a heavy-load hour.

From our model we have an estimate of u_K and α_K. The probability of a heavy-load hour in any week for an SLM system with K main stations assigned is then estimated as

$$\hat{P}(K) = 1 - \exp\left(-e^{-\hat{\alpha}_K[L(K)-\hat{u}_K]}\right). \tag{15}$$

Our goal is to let heavy-load hours occur about once every quarter of a year (13 weeks). We now invoke Gumbel's definition of *return period* as the mean time (in weeks) between heavy-load hours, which is

$$R(K) = [P(K)]^{-1}. \tag{16}$$

If we choose K so that the estimated return period is exactly 13 weeks, the temporal variation in the observed weekly peak loads will cause a distribution of return periods, among systems, centered around 13 weeks.

(Note that the expected number of heavy-load hours per year—in other words, the frequency of heavy-load hours—is $52/[R(K)]$ or $52 \cdot P(K)$, so that a 13-week return-period is equivalent to 4 heavy-load hours per year.)

If this distribution—the distribution of return periods that would be realized if K were chosen as just described, based on measured values of u_J and α_J, for each of many systems—were very narrow, so that most systems would turn out to have return periods not far from 13 weeks, then it would be appropriate to choose K in such a way as

* In the SLM system, as in many such systems, the cost of this capability would not be justified; and it would have the further disadvantage of preventing operator access to such calls.

to satisfy the equation

$$1 - \exp\left(-13e^{-\hat{\alpha}_K[L(K)-\hat{u}_K]}\right) = 0.5.$$

This would produce an even chance of having a heavy-load hour in any 13-week period. But as shown in the next section, this situation only occurs when u_J and α_J are estimated from more weeks' peak-load data than are conveniently obtainable in practice.

Furthermore, the scaling procedure described above makes K dependent on the value of J at which u_J and α_J were measured; that is, the charts corresponding to J main stations, of which Fig. 1 is the example for $J = 80$, differ considerably from each other. And this method of determining K is appreciably affected by using the chart for J working main stations (MS) when the actual number of MS served has varied widely during the measurement period, even with the correct mean of J. Thus the number of working main stations must be held nearly constant (actually within 10 percent) during the N weeks of peak-load measurements that yield $\bar{x}$ and v from eqs. (1) and (2). Yet we do not think it practical to impose limits on the growth of an SLM system's fill over periods exceeding four weeks in length, merely so as to obtain usable data, so long as the actual MS fill is not too high.

We resolve this difficulty by setting $N = 4$. This results in a very broad distribution of actual return-periods. We then locate this distribution so that its right-hand tail (in terms of MS capacities) is not too large: In particular, we limit the probability of a return period less than four weeks to a few percent. This is equivalent to setting the mean return-period for all systems equal to 37 weeks, and to having at least one heavy-load hour occur in any 13-week period with probability 0.3. Thus we choose K (to the nearest multiple of 5) so as to satisfy the equation

$$1 - \exp\left(-13e^{-\hat{\alpha}_K[L(K)-\hat{u}_K]}\right) = 0.3. \tag{17}$$

This is the basis for the charts such as that shown in Fig. 1.

The following section shows that this way of estimating K must be repeated four times, and the results averaged, in order to make the resulting distribution of return periods acceptably narrow. When this is done (to obtain what we call the *predicted* capacities), practically no systems should end up witb return periods less than four weeks—that is, with more than 1 heavy-load hour per month. This, then, is the service criterion: *No system should have more than 1 heavy-load hour in the average month.* This criterion is extremely conservative: First, it

means that the average system has less than 2 heavy-load hours per year (return period, 37 weeks). Second, it typically implies about one originating call with dial-tone delay and one blocked incoming call *per month* for the worst systems. Systems with more frequent heavy-load hours (return periods less than four weeks) would be called overloaded and would have to have main stations removed. Detection of overloaded systems by means of all-channels-busy readings is covered elsewhere. (Two heavy-load hours can, of course, occur in one month without implying an overloaded condition.) Proper use of the administrative procedure should make the occurrence of overloads extremely rare.

VI. THE MONTE CARLO STUDY

In order to evaluate the statistical variability of the main-station capacities that would be predicted by the recommended procedure, we simulated its performance in a Monte Carlo study. For each of five samples of weekly peak loads actually observed in the Rural Line Study, we estimated the parameters of their distribution from (6) and (7) and converted these, by means of (13) and (14), to estimates of the values u_J and α_J that would have existed with $J = 40$

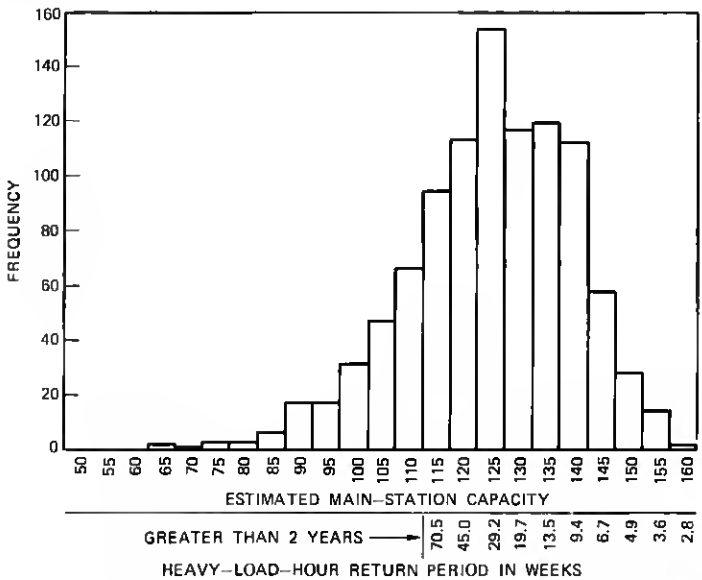

Fig. 2—Monte Carlo results: 40 working main stations; 1000 runs of 4 weeks each.

and $J = 80$ main stations working. Using these as tbe parameters of the extreme-value distribution (3), we repeatedly generated random samples of sizc four from that distribution. We calculated the mean and variance of each such sample and estimatcd tbe main-station capacity. This process was repeated 100 times for each of the five samples and for both 40 and 80 working main stations, yielding a sample distribution of the estimated main-station capacity. The spread of this distribution is attributable to the variability of the weekly peaks drawn from the distribution (3).

Figure 2 shows an example representing 1000 runs based on the 1FR* data from McComb, Mississippi, describcd in Ref. 3. Alongside tbe estimated-main-station-capacity scale is a scale which gives the mean return-period of beavy-load hours for the indicated number of main stations. Note that if the procedure were followed with only four weeks' data, some systems would be overloaded (with return periods of heavy-load hours as short as three weeks) and others, with much spare capacity, would be greatly underloaded. Table IA summarizes the Monte Carlo results for the five samples. The statistics listed are:

(*i*) The percentage of cases that would have a return period of less than the desired 13 weeks.
(*ii*) The percentage of cases that would have a return period of less than 4 weeks. These systems would be considered overloaded.
(*iii*) The percentage of cases that would be underloaded by more than 20 main stations. A system is considered underloaded if it has a return period greater than the central value of 37 weeks.

We see from Table IA that the percentage of systems that would be overloaded ranges from 2 to 14 and the percentage of systems that would be seriously underloaded ranges from 3 to 12. On the basis of these results it was decided that four weeks' data (one *measurement month*) are not enough for a final determination of main-station capacity. We therefore recommend the use of the mean of the main-station-capacity estimates of four samples of four weeks each. (This should be a weighted mean, the weights being the average numbers of working MS in the four measurement months. This weighting accounts for the greater predictive value of an estimate that is based on tbe traffic of a larger fraction of the stations that will ultimately be served.) A Monte Carlo study was carried out for this procedure and the results are shown in Table IB. Note that an overloaded or badly

* Single-party, flat-rate, residential service.

Table I — Percentage of systems that would have particularly high or low loads — summary of Monte Carlo results

	Return Period <13 Weeks		Return Period <4 Weeks		Underloaded by More Than 20 MS	
Working Main Stations:	40	80	40	80	40	80
A: ESTIMATED (100 Runs of 4 Weeks Each)						
Study Area						
Hanceville	28%	39%	3%	5%	7%	12%
Benton—1FR	21	37	2	4	7	7
Cleveland	28	38	2	14	9	7
Copper Hill	29	43	3	10	3	4
McComb—1FR	19	27	3	5	8	10
B: PREDICTED (1000 Runs of 4 Groups of 4 Weeks)						
Hanceville	11.2%	11.4%	0.0%	0.0%	0.3%	0.6%
Benton—1FR	8.8	19.7	0.0	0.0	0.1	0.5
Cleveland	10.3	22.6	0.0	0.3	0.1	0.1
Copper Hill	11.3	24.1	0.0	0.0	0.0	0.0
McComb—1FR	4.1	13.3	0.0	0.2	0.1	0.1

underloaded system would result very infrequently. Figure 3 shows a histogram of the sample distribution based on the McComb data. We distinguish the result of the modified procedure, using the mean of four estimated capacities, by calling it the *predicted* main-station capacity; and Fig. 3 is so labeled.

The last two pairs of columns in Table I relate to cases in which the administrative system may be said to have performed inadequately. The criterion for this is more stringent on the overload side, as necessitated by the inherent variability of the predicted capacities. This difficulty could be cured by taking many more measurements—a solution whose cost, in our judgment, would not be justified.

An alternative to the averaging procedure would be to calculate the mean and variance of 16 weekly peaks and to use the main-station-capacity charts only once. The precision of this approach was shown by a Monte Carlo study to be comparable to that resulting from averaging the four estimates from groups of four weekly peaks. But, as mentioned above, if a system is being filled during the measurement period, there is a much better chance of the number of working

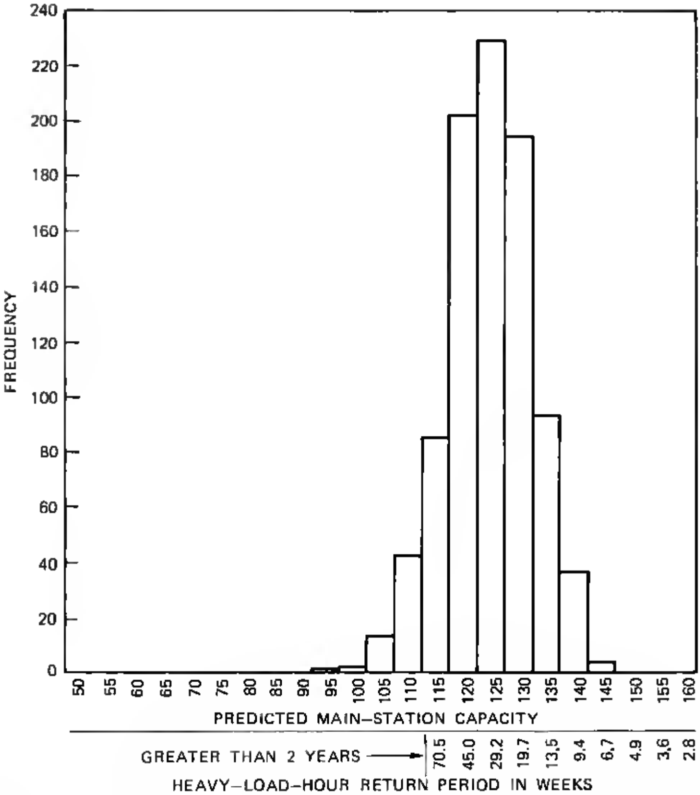

Fig. 3—Monte Carlo results: 40 working main stations; 1000 runs of 4 groups of 4 weeks.

main stations remaining nearly constant for four weeks than for 16 weeks.

The spread of the capacity estimates, as indicated by Table I, was sligbtly larger with 80 main stations working than it was with 40. We have found no convincing explanation for this unexpected result.

VII. DISCUSSION OF ASSUMPTIONS

Several assumptions used in tbe foregoing argument are known to be invalid. Nevertheless, the procedure we recommend worked well when applied to tbe Rural Line Study data. [Some confirmation of its efficacy comes from weekly peak loads observed in an SLM system trial in Brandon, Mississippi, which closely fit tbe Gumbel distribution (3).]

But in order to distinguish the realistic from the idealized features of our mathematical model, we summarize the latter in this section.

The concept of a candidate busy-hour is fictitious. However, we know of no better means of projecting the distribution of weekly peaks from the current number of working main stations to a larger number of main stations; and it seems intuitively reasonable that not all waking hours of the week should be equally likely to contain the week's highest load. Furthermore, as shown in Appendix A, the procedure does seem to be quite insensitive to the form of the assumed distribution of candidate-busy-hour loads and to the number n of candidate busy-hours. We do not believe that this aspect of our model will cause problems in practice.

It is also assumed that the loads generated during a candidate busy-hour by each of the (existing or future) working main stations are statistically independent and identically distributed. In fact, subscribers in an area do not behave independently, since they are subject to similar influences from events in their shared environment. More important is the critical assumption that the load parameters are identical for all customers. Departures from this assumption among customers already served may actually offset other errors in the model, as noted in Section V; but the possibility of adding to an SLM system a group of subscribers whose traffic characteristics differ considerably from those of the ones already being served constitutes the greatest danger in using the predictive method of administration here proposed. One way of guarding against this danger is to limit the number of main stations that may be added to a system before additional measurements are taken; and a precaution of this kind is included among the rules of administration given in Appendix B. New subscribers should be added to a system even more cautiously if they are thought to differ sociologically from those whose loads have been measured, especially if those to be added have higher incomes. Business telephones, in particular, may generate several times the loads typical of residential service.

The problem of seasonal variation has not been mentioned. No seasonal effects were observed in the Rural Line Study, but some SLM systems, especially in resort areas, will have highly seasonal loads. Our administrative procedure contains no internal safeguards against this source of error. Local knowledge will have to ensure that only weekly peak loads recorded in the busy season are used as inputs to the computations, and in some cases this may force planning to be based on fewer than four 4-week estimates of MS capacity.

In the discussion above, it has been assumed that the number of working main stations remains constant during a measurement month. The case in which inward or outward movement occurs during or between measurement months is covered by some of the rules in Appendix B.

It is not known whether any error is introduced by applying to *peak* loads the load-service relations deduced from Jones's model.

VIII. SUMMARY AND CONCLUSIONS

In this paper we describe a procedure for forecasting the main-station capacity of an SLM system when it is partially filled. The procedure is simple to use, requiring little data and few calculations. Its validity and precision have been shown to be adequate by applying it to the Rural Line Study data (as described in Appendix A) and through a Monte Carlo study of its statistical variability. An SLM system filled in the recommended way should be essentially nonblocking.

In fact, most rural systems will be able to serve many more than the nominal 80 MS. Application of our procedure to the data from the Rural Line Study led to a sample of predicted capacities whose mean slightly exceeds 120 MS, but this figure must be viewed with caution for three reasons: First, this is the mean of the limits imposed by traffic considerations alone; geographical constraints associated with transmission criteria may keep it from being attained in practice. (Lack of demand for multiparty service could also act to limit main-station fills.) Second, the Rural Line Study was conducted in eight rural areas in the territory of South Central Bell; loads would certainly be larger in some suburban applications, and as yet we have no assurance that our data are representative even of rural areas in other parts of the country. And third, the predicted capacities have a very wide spread; a small but significant fraction of systems are predicted to have a capacity of only 80 MS, confirming the early choice of that number as appropriate for rural use. Studies of suburban traffic that are now in progress should lead to an evaluation of SLM main-station capacities in the suburban environment.

In this study we have viewed the capacity of a telephone system in terms of its behavior in the presence of peak rather than average demand. (In this general sense our work has many forerunners, in the Bell System and elsewhere.) Instead of measuring the load in a time-consistent busy-hour, we record only the load carried during the busiest hour of the week. This reduces the measurements required to

only one ohservation per week hy focusing on the hour that is most important to tbe quality of service, regardless of when that bour occurs. It has the corresponding disadvantage of working witb a traffic statistic that is volatile (as compared to the more stable mean, for example) and tberefore difficult to predict witb accuracy.

Although our procedure was developed for subscriber multiplexers with concentration, we hope that modified versions of it will prove applicable to otber traffic-handling systems.

IX. ACKNOWLEDGMENT

M. M. Buchner, Jr., was deeply involved in the planning whicb led to the idea of using a peak-load measurement for the SLM system.

APPENDIX A

Number of Candidata Busy-Hours and Their Load Distribution

In the hody of this paper we assume that the candidate-busy-hour loads are normally distributed. Although the central-limit tbeorem supports this assumption, the gamma distribution has in some situations been found to be more descriptive of offered loads; and unlike the normal, it does not imply the existence of negative loads. This appendix summarizes a study comparing the normal and gamma distrihutions as bases for scaling peak loads, and also leading to tbe choice of $n = 10$ as the number of candidate busy-bours.

The gamma distrihution function, also a memher of the exponential class, takes the form

$$G(x) = \frac{1}{\Gamma(\eta)\beta^{\eta}} \int_0^x s^{\eta-1} e^{-s/\beta} ds \tag{18}$$

for $x > 0$; $G(x) = 0$ otberwise. The scale of G is determined by tbe parameter β and the shape hy η; the mean and variance are $\beta\eta$ and $\beta^2\eta$ respectively. The distribution G is asymptotically (as $\eta \to \infty$) normal. It is considerably more difficult to manipulate algehraically than is the normal: For example, such expressions as (13) and (14), which are simple for the normal, are not available for the gamma.

With a change of variahle in (18), eqs. (8) and (9) become

$$\frac{1}{\Gamma(\eta)} \int_0^{u/\beta} s^{\eta-1} e^{-s} ds = 1 - \frac{1}{n} \tag{19}$$

and

$$u\alpha = n \frac{(u/\beta)^{\eta} e^{-u/\beta}}{\Gamma(\eta)}. \tag{20}$$

The procedure requires that given n, u, and α we solve (19) and (20) for η and β. This was done by using a modified regula-falsi method of iteration. Solving for u and α when η and β are given is somewhat easier: Given η, the value of u/β can be found from a subroutine for the inverse incomplete gamma function, and from u/β the product $u\alpha$ is easily calculated from (20).

It is well-known that the sum of independent, identically distributed gamma variables is also gamma. Hence, if the candidate-busy-hour load distribution for J main stations is gamma with parameters β and η, the individual main-station mean load $\mu = \beta\eta/J$ with variance $\sigma^2 = \beta^2\eta/J$. Thus the candidate-busy-hour load distribution for K main stations is also gamma with scale parameter β and shape parameter $K\eta/J$.

We now have all that is necessary to arrive at u_K and α_K, for both the gamma and normal distributions, if we are given u_J and α_J and any value of n.

In choosing the form of the candidate-busy-hour distribution F and the number n of candidate busy-hours, we used the Rural Line Study data to evaluate the precision and bias of the estimation procedure. We chose 13 groups of lines, varying in size from 54 to 254 main stations, on which we had peak-load data in series whose lengths varied from 14 to 51 weeks. We divided each group in two in such a way that the two subgroups were approximately equal in the numbers of main stations for each class of service. We found the weekly peaks for each subgroup and, combining the two subgroups, for each whole group as well. From the subgroups' weekly peaks we predicted the distribution of the weekly peaks for each whole group and then compared these predictions with the observed whole-group weekly peaks. This was done with both normal and gamma candidate-busy-hour load distributions, for numbers of candidate busy-hours ranging from 6 to 18.

To determine the best model two measures were used: the root-mean-square deviations of the subgroup predicted values from the total-group estimated values of the mean peak load and of the 37-week-return-period load. Table II shows the results of these calculations. Since the gamma and normal assumptions perform about equally well, and since calculation of the charts (such as that shown in Fig. 1) is many times easier, faster, and cheaper with the normal, we chose it as a model for the distribution of candidate-busy-hour loads. The results of our procedure are not very sensitive to the value of n; and

Table II — Variability of predicted peak loads — square-root of the mean-squared deviation (in CCS)

Number of Candidate Busy-Hours	Mean Weekly Peak		37-Week-Return-Period Load	
	Gamma	Normal	Gamma	Normal
4	22	21	24	24
6	17	16	23	23
8	14	14	24	23
10	13	13	25	25
12	13	14	27	26
14	13	15	29	28
16	14	17	31	29
18	15	18	32	31

we took $n = 10$ because this value gave optimal performance in predicting the mean weekly peak load (the more stable statistic) and nearly optimal for the 37-week-return-period load.

Table II also gives us an evaluation of the procedure; and in particular, it tests the assumption of homogeneity within tbe groups of customers represented. The rms error for the predicted mean weekly peak is 13 CCS, or about four to six main stations. This means tbat most such predictions should be accurate to within 10 main stations. (Since these results come from 14 to 51 weeks' data, the statistical variability that would result from using only four weekly peaks is not represented here; it was investigated in the Monte Carlo study discussed in the body of this paper.)

Table III shows tbe predicted means and the whole-group sample means for the weekly peaks of the 13 groups, using the normal model with $n = 10$. There appears to be no bias in the prediction procedure. Tbe last column in Table III shows the predicted main-station capacities for SLM systems if installed in the 13 study areas. Tbese capacities range from 100 to well over 160 main stations. Excluding the five groups that are predominantly four- and eight-party subscribers, the mean main-station capacity is 124.

APPENDIX B

Rules for Traffic Administration *

In order to ensure that the conditions for validity of our mathematical model are satisfied, or nearly so, and to guard against erroneous

* The proposed procedure, together with these operational rules, is now being tested in field use. This appendix is included here only to illustrate problems that arise in reducing the peak-load approach to practice.

Table III — Comparison of predicted and observed peaks (in CCS) from the 13 Rural Line Study groups

Group Location	Number of Main Stations for Each Class of Service			No. of Weeks	Predicted Mean Weekly Peak for Total Group		Whole-Group Sample Mean Weekly Peak	Predicted Main-Station Capacity
	1FR	2FR	4 & 8FR		Subgroup I	Subgroup II		
Cullman, Ala.	49	4	33	44	230	265	246	130
Hanceville, Ala.	54	22	12	28	265	274	278	110
Jones Chapel, Ala.			254	14	413	461	426	>160
Jones Chapel, Ala.		6	114	14	324	308	308	>160
Benton, Tenn.	54			43	192	180	187	115
Benton, Tenn.		68		43	174	188	181	150
Cleveland, Tenn.	72	24		51	348	320	326	110
Copper Hill, Tenn.	71	22		40	345	323	338	100
McComb, Miss.	79			26	253	261	253	120
McComb, Miss.			205	26	461	460	477	135
Tylertown, Miss.	33	40	33	26	234	249	237	>160
Tylertown, Miss.			147	26	261	282	270	>160
Tylertown, Miss.			151	26	353	335	339	>160

predicted capacities when they are not, we give the following rules and guidelines for the use of the administrative procedure.

Data and definitions

Readings should be taken (and the WPL register reset to zero) at the same time every week. When this is not possible, each measurement week must contain no more than six weekdays and no less than four weekdays. When this condition is violated, the data should be discarded.

Measurement weeks need not be contiguous. Studies have shown that there is no serial correlation among weekly peaks. A missing week should therefore not affect the results, so long as there are four weekly peaks in each measurement month.

The number J of main stations associated with a measurement month should be the mean of the numbers of main stations being served at the times the four weekly peak loads were recorded. This mean should be rounded to the nearest multiple of ten when one of the charts such as Fig. 1 is to he chosen.

Inward or outward movement of main stations served by an SLM system will tend to increase the variance of the weekly peak loads and so to decrease thc accuracy of capacity estimates. A study has shown that if the number of main stations varies more than 10 percent during a measurement month, the peak-load data should not he used to estimate a main-station capacity.

The MS capacity predicted from several measurement months is the weighted mean of the monthly estimates. The weights are the numbers of working main stations for each measurement month.

No measurements should he used when fewer than 40 main stations are being served. For this reason, the first of the charts for estimating main-station capacities is for $J = 40$ working main stations.

Restrictions on allowad fill

No system should be loaded beyond 60 main stations until the available data (used in accordance with the rules in this appendix, and consisting of at least an estimate hased upon one measurement-month) indicate that it is safe to do so. (The Rural Line Study shows that even some rural systems may suffer excessive blocking with 80 main stations.)

No system may serve more than 160 main stations, with this exception: Unusual configurations in which no concentrator-blocking

is possihle, such as 24 lines with 8 MS on each, are perfectly acceptahle with respect to the considerations treated here.

Except for the initial fill, no more than 40 main stations may he added to an SLM system on the hasis of a single predicted capacity. If a predicted capacity exceeds the present fill by more than 40 MS, the system should be allowed to grow by 40 MS and a new series of measurements taken. This rule embodies a compromise between the value of predicting main-station capacities and the danger of adding customers unlike those already served.

If the current main-station capacity estimate was calculated from fewer than four measurement months, the numher of main stations may be increased to the lesser of

(*i*) the current main-station-capacity estimate minus 20, and
(*ii*) the mean of the current main-station-capacity estimate and the current number of working main stations.

These restrictions are necessary because of the statistical variability of such estimates.

REFERENCES

1. I. M. McNair, Jr., "Digital Multiplexer for Expanded Rural Service," Bell Laboratories Record, *50*, No. 3 (March 1972), pp. 80–86.
2. R. V. Laue, unpublished reports, 1970 and 1972.
3. J. M. Maynard, unpublished reports, 1971.
4. E. J. Gumbel, *Statistics of Extremes*, New York: Columbia University Press, 1958.
5. A. W. Jones, "Measurement of Non-Random Telephone Traffic with Special Reference to Line Concentrators," Third International Teletraffic Congress, Paper No. 21, Paris, 1961.
6. E. M. Johnson, unpublished report, 1970.
7. W. S. Hayward, Jr., "Traffic Engineering and Administration of Line Concentrators," Second International Teletraffic Congress, The Hague, 1958.